

# MASTER

BROOKHAVEN NATIONAL LABORATORY

November 1987

BNL-40512

## ALGORITHMIC ALTERNATIVES

Michael Creutz

BNL--40512

Physics Department  
Brookhaven National Laboratory,  
Upton, NY 11973

DE88 003905

### ABSTRACT

A large variety of Monte Carlo algorithms are being used for lattice gauge simulations. For purely bosonic theories, present approaches are generally adequate; nevertheless, overrelaxation techniques promise savings by a factor of about three in computer time. For fermionic fields the situation is more difficult and less clear. Algorithms which involve an extrapolation to a vanishing step size are all quite closely related. Methods which do not require such an approximation tend to require computer time which grows as the square of the volume of the system. Recent developments combining global accept/reject stages with Langevin or microcanonical updates promise to reduce this growth to  $V^{4/3}$ .

Talk given at Int. Symp. "Field Theory on the Lattice", Seillac, France 9/28-10/2/87.

---

This manuscript has been authored under contract number DE-AC02-76CH00016 with the U.S. Department of Energy. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

5-1

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## ALGORITHMIC ALTERNATIVES

Michael CREUTZ

Physics Department, Brookhaven National Laboratory, Upton, NY 11973

A large variety of Monte Carlo algorithms are being used for lattice gauge simulations. For purely bosonic theories, present approaches are generally adequate; nevertheless, overrelaxation techniques promise savings by a factor of about three in computer time. For fermionic fields the situation is more difficult and less clear. Algorithms which involve an extrapolation to a vanishing step size are all quite closely related. Methods which do not require such an approximation tend to require computer time which grows as the square of the volume of the system. Recent developments combining global accept/reject stages with Langevin or microcanonical updating<sup>s</sup> promise to reduce this growth to  $V^{4/3}$ .

### 1. INTRODUCTION

I review here the main Monte Carlo techniques in use for lattice gauge simulations. To this conference of experts in the field, most of what I say is well known. Nevertheless, it may be useful to reconsider our methods in the hope of finding efficient variations which can free computer time, thus enabling even larger scale computations than presently possible. A more extensive review of current lattice gauge algorithms is contained in Ref. 1.

I will discuss algorithms both for purely bosonic and fermionic theories. In the former case, present methods are generally quite satisfactory. Nevertheless, I will present arguments that overrelaxation can save on the order of a factor of three in such simulations.

For simulations which include the effects of quark loops, the situation is considerably more cloudy. Here the existing algorithms are extremely time consuming, perhaps requiring of order  $10^4$  times as much computer time as similar purely bosonic calculations. This places severe strains on the ability to make necessary checks. While much of this factor must inevitably come from improved machines, hopefully better algorithms can bring us there sooner.

The methods which have been proposed for fermionic simulations fall into two classes. First are

approximate approaches which involve an extrapolation in a step size parameter. Here I include the original pseudofermionic<sup>2</sup> technique as well as the microcanonical<sup>3</sup>, Langevin<sup>4</sup>, and hybrid<sup>5</sup> approaches, which involve discretization of a differential evolution. Second are those algorithms which don't require this additional extrapolation. These I refer to as "exact" methods, the prototype for which is contained in Ref. 6. Unfortunately, the latter algorithms tend to require substantially more computer time, which typically grows with the volume squared of the system.

I will argue that approaches in the first class, those involving a step size extrapolation, are all quite closely related. Indeed, they reduce the problem to an effective bosonic theory and differ primarily in the Monte Carlo algorithm applied to that system.

For the exact fermionic algorithms, the coefficient of the volume squared computer time dependence is not universal, and I will suggest that this coefficient can be reduced by a suitably biased selection of trial changes in a global set of variables. Finally, I will argue that recently proposed combinations of Langevin or microcanonical updating with a final accept/reject step use computer time growing substantially slower with the system volume.

## 2. BOSONIC FIELDS

I begin my discussion of bosonic fields with familiar trivialities. The goal of a Monte Carlo simulation is to obtain, in a stochastic manner, field configurations with a probability distribution given by the Boltzmann weight

$$P_{eq}(C) \propto e^{-S_G(C)}. \quad (1)$$

This is obtained via a Markov chain of configurations, each of which represents a pseudorandom change on the previous one.

Two obviously necessary conditions on the evolution turn out in addition to be sufficient for an approach to equilibrium. The first of these is ergodicity; that is, it should be possible for the evolution to reach any state. Indeed, if a state cannot be reached and its initial probability is zero, it will always be so. The second condition is that equilibrium be stable; that is, the algorithm should leave an equilibrium ensemble in equilibrium. The sufficiency of these two conditions is rather easily demonstrated.<sup>1</sup>

To insure the stability of equilibrium, many algorithms are formulated to satisfy a principle of detailed balance. In particular, when applied to an equilibrium ensemble, the rate of taking any configuration  $C$  to any other configuration  $\tilde{C}$  should be equal to the rate for the reverse process of taking configuration  $\tilde{C}$  to  $C$ . The former rate is just the probability  $P_{eq}(C)$  for being at  $C$  times the probability  $P(C \rightarrow \tilde{C})$  that the algorithm takes  $C$  into  $\tilde{C}$ . Explicitly, detailed balance reads

$$P(C \rightarrow \tilde{C}) P_{eq}(C) = P(\tilde{C} \rightarrow C) P_{eq}(\tilde{C}). \quad (2)$$

That this condition is sufficient for stability follows from summing over  $C$ . It is, however, not a necessary condition because one could have stability with a net circulation occurring amongst the configurations. For a trivial example, a standard microcanonical evolution follows closed orbits in phase space.

In their classic 1953 paper, Metropolis, Rosenbluth, Rosenbluth, Teller and Teller<sup>7</sup> showed that detailed balance could be easily implemented via a conditional acceptance for trial changes. In particular, consider

choosing some configuration  $\tilde{C}$  as a possible update to  $C$ . This state can be taken with an arbitrary trial probability  $P_{T,C}(\tilde{C})$ . Here I place subscript  $C$  on  $P_T$  as a reminder that the trial change can depend on the previous state. The Metropolis *et al.* scheme now accepts  $\tilde{C}$  with the conditional probability

$$P_{acc} = \min \left[ 1, \frac{P_{T,\tilde{C}}(C) e^{-S(\tilde{C})}}{P_{T,C}(\tilde{C}) e^{-S(C)}} \right]. \quad (3)$$

To check that this gives detailed balance, evaluate the left hand side of eqn. 2

$$\begin{aligned} P(C \rightarrow \tilde{C}) P_{eq}(C) &= P_{T,C}(\tilde{C}) P_{acc} e^{-S(C)} \\ &= \min \left[ P_{T,C}(\tilde{C}) e^{-S(C)}, P_{T,\tilde{C}}(C) e^{-S(\tilde{C})} \right]. \end{aligned} \quad (4)$$

Detailed balance is manifested in the explicit symmetry of this expression under exchange of  $C$  and  $\tilde{C}$ .

The general Metropolis *et al.* approach actually includes many variations depending on the choice of  $P_{T,C}(\tilde{C})$ . I will now discuss four such alternatives.

First is the conventional approach as used in most pure gauge simulations. Consider updating some link variable  $U$ , an element of the gauge group. For the trial new element  $\tilde{U}$ , multiply  $U$  by a random element  $H$

$$\tilde{U} = UH. \quad (5)$$

The element  $H$  is chosen with a probability distribution  $P(H)$  with equal probability for  $H$  or its inverse

$$P(H) = P(H^{-1}). \quad (6)$$

In practice, the matrix  $H$  is usually randomly selected from a table, and eqn. 6 is insured by having the table contain the inverse of each of its elements. Eqn. 6 implies for the trial probability

$$P_{T,U}(\tilde{U}) = P_{T,U}(U). \quad (7)$$

Thus the factors of  $P_T$  drop out of eqn. 3, giving simply

$$P_{acc} = \min \left[ 1, e^{-(\tilde{S}-S)} \right], \quad (8)$$

where  $S$  and  $\tilde{S}$  denote  $S(U)$  and  $S(\tilde{U})$ , respectively.

This approach contains an essential dependence on two parameters. The first is the average distance  $H$  lies from the identity. If this is too large, then the action change will usually be large and the trial change rarely

accepted. On the other hand, if  $H$  always lies quite close to the identity, the exploration of configuration space will be quite slow. An optimum appears to be to adjust this distance for an acceptance rate of 30-50%. The second parameter of the standard approach is the number of trial changes attempted on a given link before proceeding to the next. In most statistical problems this is taken to be unity; however, in gauge theories considerable arithmetic is needed to calculate the environment of a link, and it usually saves overall computer time to test several trial elements.

A second variation of the Metropolis *et al.* algorithm is the "heat bath" approach. Here one takes the trial element with a probability given by the Boltzmann factor with the neighboring elements held fixed

$$P_{T,U}(\tilde{U}) \propto e^{-S(\tilde{U})}. \quad (9)$$

In this case the ratio of trial probabilities

$$\frac{P_{T,\tilde{U}}(U)}{P_{T,U}(\tilde{U})} = \frac{e^{-S}}{e^{-\tilde{S}}} \quad (10)$$

exactly cancels the ratio of Boltzmann factors in eqn. 3 to give

$$P_{\text{acc}} = 1, \quad (11)$$

and the trial change is always accepted. This algorithm works quite well in those cases where the group manifold is sufficiently simple that the appropriate distribution of eqn. 9 can be easily generated. For the case of  $SU(2)$  see Ref. 8.

For a third implementation of the Metropolis *et al.* algorithm, I now discuss what is sometimes referred to as "smart Monte Carlo." In the standard approach mentioned above, the trial changes are made in an unbiased way. In the heat bath method one increases the acceptance by making trial changes with a carefully defined distribution. One might try something intermediate by combining a random change with a biasing driving force. For example, consider updating some real degree of freedom  $A$ . For a trial change take

$$\tilde{A} = A + \sqrt{\epsilon} p + \epsilon F(A). \quad (12)$$

Here  $p$  is an unbiased noise which, for the purposes of discussion, I take to be Gaussian

$$P(p) \propto e^{-p^2/2}. \quad (13)$$

The driving force  $F(A)$  is for the moment arbitrary, and  $\epsilon$  is inserted as a parameter which I will consider to be small. Solving eqn. 12 for  $p$  gives the trial probability distribution

$$P_{T,A}(\tilde{A}) \propto e^{-p^2/2}, \quad (14)$$

$$p = \frac{1}{\sqrt{\epsilon}} \left( \tilde{A} - A - \sqrt{\epsilon} F(A) \right).$$

This can be inserted into the Metropolis *et al.* acceptance criterion, eqn. 3, to give

$$P_{\text{acc}} = \min \left[ 1, \frac{\exp(-\tilde{S} - \tilde{p}^2/2)}{\exp(-S - p^2/2)} \right]. \quad (15)$$

Here  $\tilde{p}$  is the negative of the noise which would be required to move back from  $\tilde{A}$  to  $A$

$$\tilde{p} = \frac{1}{\sqrt{\epsilon}} \left( A - \tilde{A} - \sqrt{\epsilon} F(\tilde{A}) \right) \quad (16)$$

$$= p + \sqrt{\epsilon} \left( F(A) + F(\tilde{A}) \right)$$

Expanding in  $\epsilon$ , I find

$$\tilde{S} + \tilde{p}^2/2 - S - p^2/2 =$$

$$\left( \sqrt{\epsilon} p + \epsilon \left( p^2 \frac{\partial}{\partial A} + 2F(A) \right) \right) \left( \frac{\partial S(A)}{\partial A} + 2F(A) \right)$$

$$+ \mathcal{O}(\epsilon^{3/2}). \quad (17)$$

This implies that the choice

$$F(A) = -\frac{1}{2} \frac{\partial S}{\partial A} \quad (18)$$

would give

$$P_{\text{acc}} = 1 + \mathcal{O}(\epsilon^{3/2}). \quad (19)$$

Thus for small changes one can improve the acceptance over unbiased changes.

If the changes are small enough, then might consider introducing a small error by ignoring the acceptance criterion altogether and simultaneously updating all fields with eqn. 12 using the driving force in eqn. 18. At this point it is instructive to change notation and write

$$\frac{\tilde{A} - A}{\epsilon} = \frac{dA}{d\tau} \quad (20)$$

$$\frac{1}{\sqrt{\epsilon}} p = \eta.$$

Then the updating procedure reads

$$\frac{dA}{d\tau} = -\frac{1}{2} \frac{\partial S}{\partial A} + \eta \quad (21)$$

which is the standard Langevin equation. I will return later to this analysis of biased trial changes.

Finally I turn to a fourth method for picking trial changes for the Metropolis *et al.* procedure. This is based on the overrelaxation idea as discussed for Monte Carlo simulation in Ref. 9. The original implementation was only for rather special forms of the action, but this limitation has been recently overcome.<sup>10</sup> Consider updating a group element  $U$ . Let  $U_0$  denote another element of the group which approximately minimizes the action  $S(U)$  when all neighbors are held fixed. Suppose further that  $U_0$  is obtained with no explicit dependence on the element  $U$  being updated. Then a useful trial change for  $U$  is

$$\tilde{U} = U_0 U^{-1} U_0. \quad (22)$$

In some sense this element lies on the "opposite side" of  $U_0$  from  $U$ . Note that this construction satisfies the symmetry relation of eqn. 7; consequently, the acceptance of  $\tilde{U}$  follows the standard procedure of comparing a random number with the exponential of the change in the action. Note also that this is a single "hit" algorithm because multiple tries at updating a given variable will only repetitively try old elements.

I will give three motivations for this choice of trial update. First, it produces a rather large change in  $U$  while, assuming the action is reasonably symmetric around  $U_0$ , resulting in only a rather small change in the action. Thus one simultaneously obtains a high acceptance rate and a rapid flow through configuration space. Second, the value of  $U_0$  depends on neighbors which in previous sweeps have adjusted their values to accommodate  $U$ . If  $U$  is allowed to float, then these neighbors will tend to drift away and the optimum value for an updated element should lie beyond  $U_0$ . Third, recent analyses by Adler<sup>11</sup> and Neuberger<sup>12</sup> indicate that overrelaxation can help reduce the increase of correlation times as a critical point is approached.

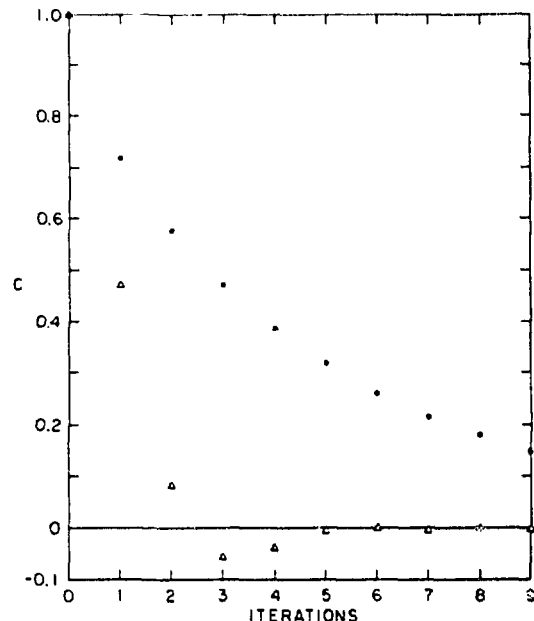


Fig. 1. The correlation between two  $SU(2)$  lattices as a function of the number of simulation iterations separating them. The model is  $SU(2)$  lattice gauge theory at  $\beta = 2.3$ . The solid points are for the heat-bath procedure while the open triangles are for the overrelaxation algorithm.

To illustrate the effectiveness of this approach I show in figure 1 some results with the gauge group  $SU(2)$ . Here I have taken an equilibrated  $7^3 \times 6$  lattice at  $\beta = 2.3$  and updated it further with both the heatbath and the overrelaxation algorithms. I plot the correlation between the links on the original lattice and those on the updated one as a function of the number of iterations or full sweeps of the lattice separating the two lattices. For the gauge group  $SU(N)$  the correlation between two lattices  $A$  and  $B$  is defined as

$$C(A, B) = \frac{1}{n_l N} \sum_l \text{ReTr}((U_l^{-1})_A (U_l)_B). \quad (23)$$

Here the sum is over all links  $l$  and  $n_l$  is the total number of links. This quantity is unity when  $A$  and  $B$  are the same, and vanishes for uncorrelated lattices.

The figure shows that while the heat-bath algorithm gives a reasonably rapid monotonic decrease of the correlation, the overrelaxation approach decorrelates still faster and in a highly damped oscillatory fashion.

In the case of  $SU(2)$  the matrix interacting with a link is always proportional to a group element. Projecting onto this element to obtain  $U_0$  gives a trial change  $\tilde{U}$  with exactly the same action, and the change will always be accepted. Thus the algorithm is deterministic and microcanonical. For  $SU(3)$  this is no longer exactly the case, but the acceptance is quite high and motion between different energy shells slow. Nevertheless, decorrelation remains quite rapid, being several times faster than optimized conventional algorithms. Recent studies of blocked correlation functions<sup>13</sup> show that the improvement in correlation times applies as well to long range gauge invariant operators.

### 3. FERMIONS

I now turn to the discussion of fermionic algorithms. As essentially all approaches begin by integrating out the fermionic fields to give a determinant, I will begin my discussion after that step and attempt to find configurations of the gauge field alone with a probability distribution given by

$$P_{eq}(A) \propto |M(A)| e^{-S_G(A)}. \quad (24)$$

I will assume at the outset that  $|M(A)|$  is a non-negative quantity, so that eqn. 24 makes sense as a probability distribution. I also will assume that  $M$  is a real matrix. This is mainly for simplicity in the following discussion, although certain complications arise with complex  $M$ .

There exists a rather straightforward scheme for updating  $A$  while satisfying detailed balance with the distribution in eqn. 24. First, just as in a Metropolis *et al.* procedure for bosonic fields, select a trial change  $\tilde{A}$  with an arbitrary trial probability  $P_{T,A}(\tilde{A})$ . Then generate a random real vector  $\chi$  with a Gaussian distribution

$$P(\chi) \propto e^{-\chi^2/2}. \quad (25)$$

Finally, accept  $\tilde{A}$  with probability

$$P_{acc} = \min \left[ 1, \frac{P_{T,\tilde{A}}(A) e^{-S_{\tilde{G}}} e^{-(\tilde{M}^{-1}M\chi)^2/2}}{P_{T,A}(\tilde{A}) e^{-S_G} e^{-\chi^2/2}} \right]. \quad (26)$$

Here  $S_G$ ,  $S_{\tilde{G}}$ ,  $M$ , and  $\tilde{M}$  denote  $S_G(A)$ ,  $S_G(\tilde{A})$ ,  $M(A)$ , and  $M(\tilde{A})$ , respectively. Note the similarity of this with eqn. 3 for the bosonic case; only the final factor is new.

To justify this procedure, I verify detailed balance. As  $\chi$  is determined stochastically, this involves an integral over that field. To study the rate of flow of  $A$  into  $\tilde{A}$ , consider

$$P(A \rightarrow \tilde{A}) P_{eq}(A) \propto \int d\chi e^{-\chi^2/2} P_{T,A}(\tilde{A}) P_{acc} P_{eq}(A). \quad (27)$$

Now change variables of integration from  $\chi$  to  $\phi = M\chi$ , noting that the Jacobian for this transformation is contained in the factor of  $P_{eq}(A)$ . This gives

$$P(A \rightarrow \tilde{A}) P_{eq}(A) \propto \int d\phi \min \left[ P_{T,A}(\tilde{A}) e^{-S_G(A)} e^{-(M^{-1}\phi)^2/2}, (A \leftrightarrow \tilde{A}) \right]. \quad (28)$$

As with eqn. 4, detailed balance is manifested in the symmetry of this expression under interchange of primed and nonprimed indices.

Let me now make three remarks on this procedure. First, this is a disguised form of the algorithm presented by Weingarten and Petcher.<sup>6</sup> Indeed, the coupled distribution for the fields  $A$  and  $\phi = M\chi$  is

$$P(A, \phi) \propto e^{-S_G - (M^{-1}\phi)^2/2}. \quad (29)$$

Second, I note that a slight variation of the algorithm makes the replacement

$$\tilde{M}^{-1}M\chi \rightarrow M\tilde{M}^{-1}\chi \quad (30)$$

in the acceptance criterion. While the factors shift around a bit, it is still easy to verify detailed balance. This variation is the "look ahead" algorithm discussed by Grady.<sup>14</sup>

Finally, note that to construct  $\tilde{M}^{-1}M\chi$  requires the inverse of our large fermionic matrix  $\tilde{M}$  applied to a given vector. This requires the solution of a large system of linear equations via a conjugate gradient or other procedure. As a consequence the algorithm requires computer time growing as the volume of the system squared. One factor of volume comes from

the sweep over the system variables, and another from the optimistic assumption that the conjugate gradient algorithm will adequately converge in a fixed number of iterations, each of which requires a time growing linearly with the volume.

This time growth as the volume squared is common to many exact algorithms, and is considerably better than the naive direct calculation of the determinant of  $M$ , which would give a fourth power of the volume behavior. In this and the following discussion of volume dependences, I assume that all couplings are being held fixed and ignore any additional slowing from critical behavior.

Because of the severe computational requirements with this algorithm, most current simulations make an additional approximation involving a small step size. The hope is to avoid calculating  $\tilde{M}^{-1}M\chi$  after each change by assuming that small changes will allow but a single conjugate gradient inversion per sweep of the entire lattice. In particular, if  $\tilde{M}$  is near  $M$ , then

$$\begin{aligned} \tilde{M}^{-1} &= M^{-1} - M^{-1}(\tilde{M} - M)M^{-1} \\ &+ \mathcal{O}\left((\tilde{M} - M)^2\right). \end{aligned} \quad (31)$$

This implies

$$\frac{e^{-(\tilde{M}^{-1}M\chi)^2/2}}{e^{-\chi^2/2}} = \frac{e^{\xi\tilde{M}\chi}}{e^{\xi M\chi}} + \mathcal{O}\left((\tilde{M} - M)^2\right) \quad (32)$$

where

$$\xi = M^{\dagger-1}\chi. \quad (33)$$

Ignoring the higher order corrections, this means one can do an ordinary bosonic updating of the  $A$  fields while holding fixed  $\xi$  and  $\chi$  in the effective action

$$S_T = S_G - \xi M\chi. \quad (34)$$

At the end of any calculation one should extrapolate observables to a vanishing value for the size of the trial changes.

This framework forms the basis of several popular fermionic algorithms. R. Gavai and I<sup>15</sup> proposed using the above effective action in a conventional Metropolis *et al.* scheme. Other approaches being pursued are Langevin updating,<sup>4</sup> microcanonical methods<sup>3</sup> and

hybrids of these.<sup>5</sup> The point I wish to emphasize here is that these algorithms all treat the fermionic part of the problem in the same manner; they differ only in the approach used to simulate the bosonic problem defined by the action in eqn. 34.

The pseudofermionic method<sup>2</sup> differs slightly from the above. Here instead of a single pair  $(\chi, \xi)$  of auxiliary bosonic fields, an ensemble of such is taken. In particular, consider fields  $\chi_i$ ,  $i = 1 \dots N$ , each chosen with Gaussian distribution  $P(\chi_i) \propto e^{-\chi_i^2/2}$ , and fields  $\xi_i = M^{\dagger-1}\chi_i$ . The pseudofermionic scheme is equivalent to updating the  $A$  fields with a small step size using an average of eqn. 34 over this ensemble

$$S_T \rightarrow S_G - \frac{1}{N} \sum_i \xi_i M\chi_i. \quad (35)$$

To justify this, elementary manipulations show that as  $N \rightarrow \infty$

$$\frac{\exp\left(\frac{1}{N} \sum \xi \tilde{M}\chi\right)}{\exp\left(\frac{1}{N} \sum \xi M\chi\right)} \rightarrow \frac{|\tilde{M}|}{|M|} + \mathcal{O}\left((\tilde{M} - M)^2\right). \quad (36)$$

Thus the method directly calculates changes in an effective action including the fermionic determinant. Ref. 16 showed two distinct ways to remove the above higher order corrections to give exact, but volume squared, algorithms also based on an ensemble average over pseudofermionic fields.

As the number  $N$  of pseudofermionic fields decreases from a large number to unity, one interpolates between this algorithm and the previously mentioned approaches. Figure 2, taken from Ref. 15, shows the variation of the average plaquette with step size for various numbers of pseudofermionic fields.

In the original presentation of the pseudofermionic algorithm a Monte Carlo technique was used to obtain the auxiliary fields, rather than the above method using conjugate gradient inversions on a random field. The field  $\xi = M^{\dagger-1}\chi$  has probability distribution

$$P(\xi) \propto e^{-(M^{\dagger}\xi)^2/2}. \quad (37)$$

As this function involves only local interactions of the  $\xi$  field, it is easily simulated by standard Monte Carlo simulation. Indeed, in all the algorithms discussed above, one has two options. One could generate  $\chi$  with Gaussian random numbers and then find

$\xi = M^{-1}\chi$  via a time consuming conjugate gradient inversion. Alternatively, one could to a time consuming Monte Carlo to obtain  $\xi$  directly with the distribution in eqn. 37 and then calculate  $\chi = M^T\xi$ . Which of these options is preferable may be only a matter of taste.

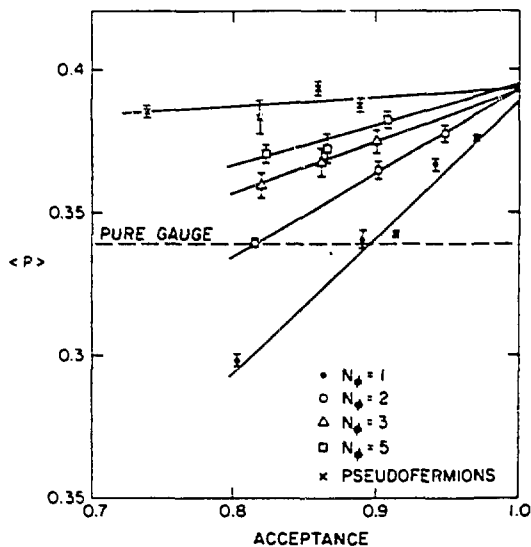


Fig. 2. The average plaquette as a function of the acceptance probability for a Metropolis *et al.* hit. The various values of  $N_\phi$  represent the number of pseudofermionic fields averaged in the effective action. The model is  $SU(3)$  with 8 flavors on a  $4^4$  site lattice at  $\beta = 4.5$ . All curves should extrapolate to the same value at unit acceptance. The lines are linear fits.

I now return to exact algorithms and discuss a class of such wherein a large number of variables are updated simultaneously via a global accept/reject step. The motivation for this is that the exact new action is hard to calculate, requiring time proportional to the volume even when only a single variable is changed, and thus it might be desirable to calculate the action less often by updating many variables simultaneously. This forms the basis for algorithms discussed in Refs. 17-19. The idea, however, has a danger that one will have large

increases in the action if many variables are changed and thus final acceptances may be unfeasibly small.

To see the potential problem, first consider making an unbiased trial change on some variable  $A$  and pick

$$\bar{A} = A + \delta A \quad (38)$$

where, because the change is unbiased,

$$\langle \delta A \rangle = 0. \quad (39)$$

A simple Metropolis *et al.* algorithm would accept this with probability

$$P_{\text{acc}} = \min \left[ 1, e^{-(\bar{S}-S)} \right]. \quad (40)$$

To analyze this further, consider the expected value of the action change

$$\langle \bar{S} - S \rangle = \left\langle \delta A \frac{\partial S}{\partial A} + \frac{\delta A^2}{2} \frac{\partial^2 S}{\partial A^2} \right\rangle + \mathcal{O}(\delta A^3). \quad (41)$$

Here the expectation value is both over the trial changes and over equilibrium  $A$  fields with  $P(A) \propto e^{-S}$ . The first term on the right hand side of eqn. 41 vanishes by virtue of eqn. 39. For the remaining term, use the identity

$$\left\langle \frac{\partial^2 S}{\partial A^2} \right\rangle = \left\langle \left( \frac{\partial S}{\partial A} \right)^2 \right\rangle, \quad (42)$$

which follows from a partial integration in the definition of expectation value at equilibrium. Ignoring the higher order terms gives

$$\langle \bar{S} - S \rangle = \left\langle \frac{\delta A^2}{2} \right\rangle \left\langle \left( \frac{\partial S}{\partial A} \right)^2 \right\rangle \geq 0. \quad (43)$$

Thus I conclude that on the average a random change increases the action by an amount proportional to the square of the step size.

Now consider updating a large number  $V$  of independent variables together. If they are all given simultaneous unbiased trial changes, then there will be a coherent addition of the above average trend. Thus the expectation for the overall acceptance is

$$P_{\text{acc}} \sim e^{-CV(\delta A)^2} \quad (44)$$

where  $C$  is a constant factor. To maintain a reasonable acceptance one should take  $\delta A \sim 1/\sqrt{V}$ . Now suppose we make  $N$  sweeps over our lattice. The total

distance  $\Delta A$  that a given variable changes should be approximately given by the random walk formula

$$\Delta A \sim \delta A \sqrt{N} \sim \sqrt{N/V}, \quad (45)$$

and the number of sweeps required to obtain a substantially different configuration should grow as  $V$ . If  $V$  is proportional to the system volume, then the overall algorithm requires time growing as volume squared, one factor of volume from the number of sweeps, and the other from the fact that each sweep is proportional to the volume.

For a bosonic simulation this growth would be a disaster. The conventional algorithms only grow as the system volume, and thus any gain involving updating many variables simultaneously would require some major additional cleverness. However, for exact fermionic algorithms we already are starting with a volume squared behavior, and thus nothing appears to be lost by going to global updates.

I now generalize the above discussion to biased global changes. To proceed, recall the treatment in eqns. 12-19. Thus consider the trial change  $\tilde{A} = A + \sqrt{\epsilon} p + \epsilon F(A)$  and acceptance  $P_{\text{acc}} = \min[1, e^{-\alpha}]$  where

$$\alpha = S(\tilde{A}) + \bar{p}^2/2 - S(A) - p^2/2 \quad (46)$$

with  $\bar{p} = p + \sqrt{\epsilon} (F(A) + F(\tilde{A}))$ . Using  $\langle p^2 \rangle = 1$  and the vanishing expectation for odd powers of  $p$  gives

$$\begin{aligned} \langle \alpha \rangle = \\ \epsilon \left\langle \left( \frac{\partial}{\partial A} + 2F(A) \right) \left( \frac{\partial S(A)}{\partial A} + 2F(A) \right) \right\rangle \quad (47) \\ + \mathcal{O}(\epsilon^3). \end{aligned}$$

A partial integration similar to that used for eqn. 42 converts this to

$$\langle \alpha \rangle = \epsilon \left\langle \left( \frac{\partial S(A)}{\partial A} + 2F(A) \right)^2 \right\rangle + \mathcal{O}(\epsilon^3). \quad (48)$$

Thus again a negative quantity appears in the exponent in the acceptance condition. I will return shortly to the case where  $F$  is chosen as in eqn. 18 so that the  $\mathcal{O}(\epsilon)$  term vanishes.

With  $V$  variables being updated simultaneously, eqn. 44 is now replaced by

$$P_{\text{acc}} \sim e^{-CV\epsilon}. \quad (49)$$

To have reasonable acceptance requires  $\epsilon \sim 1/V$ . Since  $A$  now has a driven drift, eqn. 45 changes to

$$\Delta A \sim \epsilon N \sim N/V. \quad (50)$$

As before  $N$  must grow as the system volume and the total computer time grows as the volume squared. However, now the coefficient of this volume squared behavior depends on the choice of the driving force  $F(A)$ . In particular, it may be possible to reduce this coefficient to a practical size.

One way to do this would be to update the entire lattice with a simply calculable effective action  $S_{\text{eff}}$ . For example, one might begin by merely shifting the gauge coupling in the pure gauge action, and later use an effective action motivated by the hopping parameter expansion.<sup>20</sup> This sweep should generate a trial new configuration with a probability satisfying detailed balance with respect to the effective action

$$P_{T,C}(\tilde{C}) e^{-S_{\text{eff}}} = P_{T,\tilde{C}}(C) e^{-\tilde{S}_{\text{eff}}}. \quad (51)$$

Finally, to make the algorithm exact, accept  $\tilde{C}$  with conditional probability

$$P_{\text{acc}} = \min \left[ 1, \frac{e^{-S_{\text{eff}}} e^{-\tilde{S}}}{e^{-\tilde{S}_{\text{eff}}} e^{-S}} \right] \quad (52)$$

where  $S$  is the full action including fermionic effects. For example, using the formalism of Ref. 6, this would be

$$S(A) = S_G + (M^{-1}(A)\phi)^2/2 \quad (53)$$

Given such a procedure, one would adjust both the step size and the parameters in  $S_{\text{eff}}$  to optimise decorrelation times.

This essentially concludes what I actually presented in my talk at Seillac. Let me now go on and discuss further what happens when the Langevin choice  $F(A) = -\frac{\partial S}{\partial \bar{A}}$  is made for the driving force in the trial updating. In this case the order  $\epsilon$  term in eqn. 48 vanishes; thus, it is desirable to consider higher orders. Remarkably, identities similar to eqn. 42 yield an expected value for the order  $\epsilon^2$  term which also vanishes.

Going still one order further, and using more partial integrations, I find

$$\langle \alpha \rangle = \epsilon^3 \left\langle \frac{5}{96} \left( \frac{\partial^3 S(A)}{\partial A^3} \right)^2 + \frac{1}{32} \left( \frac{\partial^2 S(A)}{\partial A^2} \right)^3 \right\rangle + \mathcal{O}(\epsilon^4). \quad (54)$$

Assuming this expectation is positive, which I have not proven in general, the above argument when  $V$  variables are updated simultaneously gives  $\epsilon \sim 1/V^{-1/3}$  and an overall time for decorrelation

$$T \sim V^{4/3}, \quad (55)$$

a growth only slightly worse than the linear growth of the pure bosonic theory.

This algorithm was proposed in Ref. 17 and tested further with somewhat discouraging results in Ref. 18. Ref. 19, discussed by Kennedy at this conference, presents an interesting variation on the above. I will now discuss a slight variation on that variation. With biased changes, one constructs both the trial new  $A$  and the negative of the reverse noise needed to return

$$\begin{aligned} \bar{A} &= A + \sqrt{\epsilon} p + \epsilon F(A) \\ \bar{p} &= p + \sqrt{\epsilon} \left( F(A) + F(\bar{A}) \right). \end{aligned} \quad (56)$$

If  $A$  is initially in equilibrium, then the coupled probability for  $A$  and  $p$  is

$$P(A, p) \propto e^{-S-p^2/2}. \quad (57)$$

If we consider the conditional acceptance as being for both  $\bar{A}$  and  $\bar{p}$  considered as a pair, then this coupled probability will be preserved.

The set of eqns. (56) represents a discretization of a microcanonical evolution of  $A$  and  $p$  under a canonical Hamiltonian  $H = S(A) + p^2/2$ . The microcanonical time step is  $\sqrt{\epsilon}$ . Furthermore, this particular mapping exactly preserves areas in phase space

$$dA dp = d\bar{A} d\bar{p}. \quad (58)$$

The scheme proposed in Ref. 19 is essentially to iterate eqn. 56 several times before making the accept/reject decision. This makes the hybrid algorithm of Ref. 5 exact, just as the procedure with a single step makes

the Langevin evolution exact. The hope is that the microcanonical evolution will sufficiently restrict changes in the action that the final acceptance will be remain high even for reasonable step sizes.

This procedure contains several parameters which can be adjusted for optimization. First is the number of iterations of eqn. 56 before the global accept/reject step followed by a refreshing of the momenta  $p$ . This corresponds directly to the refreshing frequency in Ref. 5. Then there is the step size  $\epsilon$ , which presumably should be set to give an acceptance in the range 30-50%. Finally, one can also vary the frequency with which the auxiliary scalar fields  $\phi$  are updated. The preliminary results in Ref. 19 are quite encouraging, and more studies are needed.

#### ACKNOWLEDGMENT

Work supported by U.S. Department of Energy under Contract No. DE-AC02-76CH00016.

#### REFERENCES

1. M. Creutz, in *Lattice Gauge Theory Using Parallel Processors*, Xiaoyuan Li, Zhaoming Qiu, and Haicang Ren, eds., p. 105 (Gordon and Breach, 1987).
2. F. Fucito, E. Marinari, G. Parisi and C. Rebbi, *Nucl. Phys.* **B180**[FS2], 369 (1981).
3. J. Polonyi and H.W. Wyld, *Phys. Rev. Lett.* **51**, 2257 (1983); J. Kogut, J. Polonyi, H.W. Wyld and D.K. Sinclair, *Phys. Rev. Lett.* **54**, 1475 (1985).
4. G.G. Batrouni, G.R. Katz, A.S. Kronfeld, G.P. Lepage, B. Svetitsky and K.G. Wilson, *Phys. Rev.* **D32**, 2736 (1985); A. Ukawa and M. Fukugita, *Phys. Rev. Lett.* **55**, 1854 (1985).
5. S. Duane and J. Kogut, *Phys. Rev. Lett.* **55**, 2774 (1985); *Nucl. Phys.* **B275**, 398 (1986).
6. D. Weingarten and D. Petcher, *Phys. Lett.* **99B**, 333 (1981).
7. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
8. M. Creutz, *Phys. Rev.* **D21**, 2308 (1980).
9. S. L. Adler, *Phys. Rev.* **D23**, 2901 (1981); C. Whitmer, *Phys. Rev.* **D29**, 306 (1984); Princeton University PhD thesis (1984).
10. F. R. Brown and T. J. Woch, *Phys. Rev. Lett.* **58**, 2394 (1987); M. Creutz, *Phys. Rev.* **D36**, 515 (1987).

11. S. L. Adler, Institute for Advanced Study preprint IASSNA-HEP-87/27 (1987).
12. H. Neuberger, Rutgers University preprint RU-35-87 (1987).
13. R. Gupta, G. Kilcup, and S. Sharpe, private communication..
14. M. Grady, Phys. Rev. D32, 1496 (1985).
15. M. Creutz and R. Gavaì, Nucl. Phys. B280, 181 (1987).
16. G. Bhanot, U.M. Heller and I.O. Stamatescu, Phys. Lett. 129B, 440 (1983).
17. R. T. Scalettar, D.J. Scalapino and R.L. Sugar, Phys. Rev. B34, 7911 (1986).
18. S. Gottlieb, W. Liu, D. Toussaint and R.L. Sugar, Phys. Rev. D35, 2611 (1986).
19. S. Duane, A. D. Kennedy, B.J. Pendleton and D. Roweth, Florida State University preprint FSU-SCRI-87-27 (1987).
20. A. Hasenfratz and P. Hasenfratz, Phys. Lett. 104B, 489 (1981).